# Navigating a Fine Balance: Point-Mutant Cheater Viruses Disrupt the Viral Replication Cycle

Moran Meir [1], Arielle Kahn [1], Carmel Farage [1], Yael Maoz [1,2], Noam Harel [1,2], Adi Ben Zvi [1,2], Shir Segev [1], Maria Volkov [1], Ravit Yahud [1], Uri Gophna [1,2], Adi Stern [1,2,*]

[1]The Shmunis School of Biomedicine and Cancer Research, Tel Aviv University, Tel Aviv, Israel

[2]Edmond J. Safra Center for Bioinformatics, Tel Aviv University, Tel Aviv, Israel

***Corresponding author:** E-mail: sternadi@tauex.tau.ac.il.

**Associate editor:** Gonzalo Moratorio

## Abstract

**Cheater viruses cannot replicate on their own yet replicate faster than the wild type (WT) when the 2 viruses coinfect the same cell. Cheaters must possess dual genetic features: a defect, which leads to their inability to infect cells on their own, and a selective advantage over WT during coinfection. Previously, we have discovered 2 point-mutant cheaters of the MS2 bacteriophage. Here, we set out to discover the possible repertoire of cheater MS2 viruses by performing experimental evolution at a very high multiplicity of infection. Our results revealed a third point-mutant cheater that arose in 8 biological replicas. Each of the 3 primary cheaters disrupts the fine balance necessary for phage replication, in different ways that create a defect + advantage. We found that over time, the point-mutant cheaters accumulate additional secondary mutations, which alter other stages of the viral replication cycle, complementing the disruptions created by the original cheater. Intriguingly, cheater and secondary mutations almost always reside in very close proximity on the genome. This region encodes for multiple functions: overlapping reading frames as well as overlapping RNA structures critical for transitioning from one stage to another in the viral replication cycle. This region of overlap explains the dual functions of cheaters, as one mutation can have pleiotropic effects. Overall, these findings underscore how viruses, whose dense genomes often have overlapping functions, can easily evolve point-mutant cheaters, and how cheaters can evolve to alter the intricate balance of the viral replication cycle.**

***Key words:*** **sociovirology**, **defective interfering particles**, **virus evolution**, **multilevel selection**.

## Introduction

Viruses are ultimate parasites that rely on cellular infection to complete their replication cycle. In recent years, there is an increasing interest in what happens when 2 or more viruses infect the same cell. In some cases, coinfection may lead to recombination or reassortment, processes that have a dramatic impact on the evolution of many viruses such as HIV and Influenza (Gao et al. 2011; Rith et al. 2015; Wu et al. 2024). Alternatively, coinfection may occur between similar variants of the same virus, leading to an ecological interaction between the 2 coinfecting viruses, which can be classified as cooperative or antagonistic (Segredo-Otero and Sanjuán 2022), concepts explored within the emerging field of "sociovirology" (DaPalma et al. 2010; Altan-Bonnet and Chen 2015; Díaz-Muñoz et al. 2017; Santiana et al. 2018; Sanjuán and Thoulouze 2019).

Defective virus genomes (DVGs) that can thrive in the presence of coinfecting wild type (WT) viruses, also known as defective interfering particles or cheater viruses are attracting increasing interest. As we will go on later to define in the context of this paper, we will mainly use the term cheater viruses. Often, cheater virus genomes are truncated, lack essential genes, and are structurally distinct from the WT genomes. They rely on the coinfecting WT genomes to supply their gene products, also denoted as "public goods" (Ciota et al. 2012; Chao and Elena 2017; Díaz-Muñoz et al. 2017); they furthermore subvert resources from the WT genomes, thus negatively impacting the replicative capacity of WT. Recent interest in cheater viruses has grown due to their potential therapeutic applications and a growing understanding on their potential effects on evolutionary processes.

On the therapeutic side, by competing for resources with pathogenic WT viruses, cheater viruses can reduce the infectious load and stimulate both adaptive and innate immune responses (Dimmock and Easton 2014; Rast et al. 2016; Xu et al. 2017; Bdeir et al. 2019; Genoyer and López 2019; Tanner et al. 2019; Leeks et al. 2023a; Brennan and Sun 2024). There is a growing interest in the use of therapeutic interfering particles that are effectively cheater viruses (Chaturvedi et al. 2021; Pelz et al. 2021; Rezelj et al. 2021), and this has stimulated increased research in the field.

**Received:** September 18, 2024. **Revised:** December 03, 2024. **Accepted:** December 08, 2024

Viral cheating also has an impact on both short-term and long-term evolutionary dynamics. First, natural selection may be reduced in the presence of cheaters through the presence of defective cheaters or defective genomes that may thrive during coinfection. Cheaters may also lead to a collapse in the viral population, with implications on effective viral population size and weaker selection (Leeks et al. 2021; Singhal and Turner 2021). Cheater viruses may impose selection on the WT viruses to avoid invasion by cheaters, e.g. via superinfection exclusion (Hunter and Fusco 2022), and may even impose selection for new genomic forms, as has been suggested in the case of multipartite viruses (Leeks et al. 2023b). Finally, cheaters add to the viral population an additional layer of genetic diversity that may lead to phenotypic novelty, similar to other selfish elements (e.g. transposons) that sometimes undergo exaptation and evolve to take on new functions (Hall et al. 2020; Fueyo et al. 2022).

We have previously discovered 2 point-mutant cheater viruses during experimental evolution of the MS2 bacteriophage, highlighting that cheating can occur not only via large genomic and structural variations, but via a mere point mutation (Meir et al. 2020). We hereby define a cheater virus as a virus bearing 2 traits: (i) a defect that reduces its ability to replicate on its own as compared to WT viruses, and (ii) a fitness advantage over WT during coinfection of cells with both WT and the cheater virus (Leeks et al. 2021). The second condition is similar to a fitness cost imposed on the WT (Ghoul et al. 2014); yet, the definition herein allows to focus only on cheater characteristics.

MS2 is a model virus that has been studied in depth, due to its applications in biotechnology and the pioneering insights it has allowed into genome function and the viral replication cycle (Dang et al. 2023; Naskalska and Heddle 2024). Its replication cycle consists of 4 main stages: (I) entry, (II) replication, (III) packaging, and (IV) lysis/exit (Fig. 1a) (Duin and Tsareva 2006; Rolfsson et al. 2016). Each of these 4 stages is controlled by 1 of the 4 proteins the virus encodes for, respectively: (I) maturation (also known as A-protein), (II) replicase, (III) coat, and (IV) lysis (Fig. 1b). Transitions between the 4 stages of the viral life cycle are tightly controlled by RNA structures in the virus genome that overlap the viral open reading frames (ORFs). The long-range Min-Jou (MJ) structure impacts replicase translation and hence genome replication (Van Himbergen et al. 1993). It has further been shown that the upstream coat terminator (CT) structure, residing immediately upstream the MJ, also impacts genome replication. Binding of coat protein to the TR loop structure initiates packaging of the virus (Johansson et al. 1997). Expression of lysis protein, and hence lysis of cells and exit of viruses, is controlled by the lysis hairpin (LH) loop (Fig. 1, supplementary fig. S1, Supplementary Material online) (Van Himbergen et al. 1993; Betancourt 2009).

Briefly, the 2 MS2 point-mutant cheaters we previously found (Meir et al. 2020) consisted of a point deletion (Δ1764) and a synonymous mutation (A1664G) that impacts the LH structure. We were able to show that

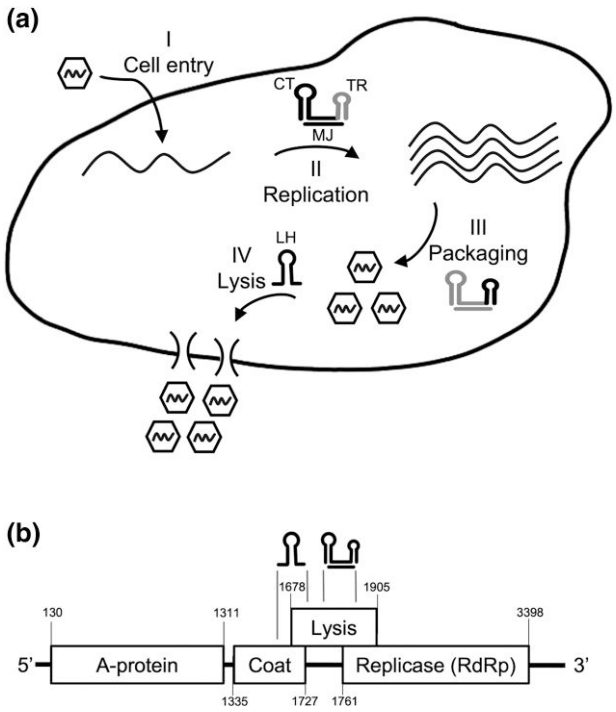


**Fig. 1.** Scheme of MS2 replication cycle and genome. a) A simplified schematic of the replication cycle of MS2 that is composed of 4 main stages: (I) Viral entry, which includes attachment to the host cell, and entry of viral RNA into the host cell, (II) Replication of positive sense RNA genomes (via negative strand intermediates—not shown), (III) Packaging and assembly of viral genomes into virions, and (IV) Lysis of cells and exit of virions. All 4 stages are mediated through translation of a relevant viral protein and recruitment of host factors as well (not shown). Transitions between different stages are mediated through the illustrated viral RNA structures (main text). As the CT, MJ, and TR loop structures are adjacent, the relevant structure is highlighted in black versus gray (stages II and III). b) A schematic of the MS2 genome with its 4 ORFs, each of which mediates a different stage of the replication cycle (main text). RNA secondary structures are illustrated above the genome. The MJ structure is a long-range structure that spans beyond the boundaries of the illustration.

Δ1764 was deleterious, therefore unable to replicate on its own, yet bore an advantage during coinfection via enhanced RNA genome packaging. Conversely, A1664G was defective on its own due to reduced lysis protein translation (producing tiny plaques), yet bore an advantage during coinfection likely via enhanced packaging as well. We denoted the phenotype of Δ1764 as full cheating since its fitness was zero on its own, and the phenotype of A1664G as semicheating, since its fitness was higher than zero on its own.

In this work, we set out to more systematically characterize the repertoire of cheater MS2 viruses and their evolution. A key term that impacts the ability of cheaters to emerge and thrive is that of multiplicity of infection (MOI), the ratio between the number of infectious viruses as assessed by plaque assay and the number of bacterial cells. We focused first on a very high MOI of 10, where coinfections occur at a high rate during each passage. We further explored longer-term evolution of our previously

discovered cheaters that evolved at an MOI of 1. Notably, when cheaters emerge, they lead to an effective increase in the MOI (they are not counted during plaque assay that is used to determine MOI); thus, the effective MOI increases as cheater frequencies increase. It is intriguing to note that the cheaters we originally found emerged despite an initially low MOI of 1 that allows for coinfections at an arguably low rate.

One of our goals was to test whether cheaters can correct their defect via compensatory mutations that corrected the defect. We found that different cheaters found in both experiments gained specific sets of additional secondary mutations on their genomes; yet, these did not correct the defect of the primary cheater mutations. We tested the impact of these secondary mutations and found that they tend to alter another 1 of the 4 stages of the replication cycle, one not necessarily overlapping that of the defect. We revealed that both cheater mutations and secondary mutations tend to reside in a very specific region of the genome with multiple overlapping functions. We propose a "fine balance model" whereby cheaters occur at regions of overlapping functions, disrupt 1 stage of the replication cycle and then evolve to alter a different stage, while overall maintaining high fitness during coinfection.

## Results

### Parallel Evolution During Serial Passaging at MOI of 10

To test if additional point-mutant cheaters emerge under conditions of higher MOI, we performed serial passaging at an MOI of 10. Passaging was performed across 3 different founder populations, resulting in 8 biological replicas. The large number of replicas was performed to ensure that founding populations do not affect the biological outcome during passaging (Methods). Strikingly, we observed parallelism across all of the 8 replicas and revealed that the same set of 5 mutations increased in frequency in parallel across all replicas (Fig. 2a), a strong signal of positive selection. The highest frequency was obtained for mutation A1744G, reaching a frequency of up to 75%. For simplicity, we will denote this mutation as A1744G-pink (Fig. 2). One of the 5 mutations observed was A1664G, the semicheater we have previously reported, which we henceforth refer to as A1664G-orange. Three additional mutations were also repeatedly observed: G1688U, C1735U, and U1829C, which we will denote as green secondary mutations.

We next used synthetic long-read sequencing (Methods) to determine which of the 5 mutations we discovered reside on the same genome. The results showed a very consistent pattern: A1744G-pink was either on its own, or with exactly 1 additional green secondary mutation (G1688U, C1735U, or U1829C). These latter green secondary mutations were almost never found on their own, were never together with one another on the same genome and were never found with the semicheater A1664G-orange (Fig. 2b, supplementary fig. S2, Supplementary Material online).

To further explore the repertoire of cheating in MS2, we used the long-read sequencing data to search for genomes with large deletions, a hallmark of cheaters in other viruses (Secor and Dandekar 2020; Pelz et al. 2021; Wu et al. 2022). We noted the presence of a variety of shorter genomic fragments with unknown biological significance; particularly, genome fragments around 1,800 nucleotides (supplementary fig. S3, Supplementary Material online). However, their frequency reached at most 0.5%, suggesting that they are not potent cheaters with a large advantage. Moreover, we did not detect enrichment or depletion of the point-mutant cheaters described below on the background of these shorter genomes.

### A Point-mutant Cheater A1744G-Pink is Inferred

We set out to test whether the newly discovered mutations A1744G-pink and the green secondary mutations are simply bona fide adaptive mutations or represent cheater viruses. To this end, we performed 3 additional passages using 4 passage 10 populations (lines A, E, G, H), at an MOI of 0.01. These conditions do not allow for coinfection and thus allow propagation of bona fide adaptive mutations but not of cheaters. Our results showed that A1744G-pink and the green secondary mutations decreased to a frequency of zero. On the other hand, we observed an increase in the frequency of A535G-turquoise (Fig. 2c). This is a mutation we have previously shown to be a bona fide adaptive mutation (Meir et al. 2020), supported by the results herein. A535G-turquoise creates a nonsynonymous mutation in the A-protein, thus likely allowing more rapid entry into cells.

We also sequenced a set of 32 plaques from a late passage and included plaques of various sizes (Methods). Our aim was to test if our newly discovered cheaters are full cheaters or semicheaters. Full cheaters are incapable of forming plaques independently due to their reliance on coinfection, while semicheaters can create smaller plaques as previously shown (Meir et al. 2020). We did not observe the A1744G-pink and green secondary mutations in any of the 32 isolated plaques (supplementary fig. S4, Supplementary Material online). Altogether, our results suggest that A1744G-pink is the driver mutation here that serves as a full point-cheater mutation. We propose that the green secondary mutations, which are observed in all 8/8 replicas, provide some type of added benefit to the cheater, as explored in more detail in the upcoming paragraphs (Table 1).

### Mechanism of Pink Cheating and Corrections by Green Secondary Mutations

Per our definition, cheaters bear both a defect during singular infection, as well as an advantage during coinfection. We set out to infer these 2 traits for A1744G-pink. This was challenging since this mutation creates 3 potential phenotypes (Table 1): (i) a nonsynonymous mutation at the lysis protein (K23E), previously shown to negatively impact lysis of cells (Chamakura et al. 2017), (ii) alteration of the MJ

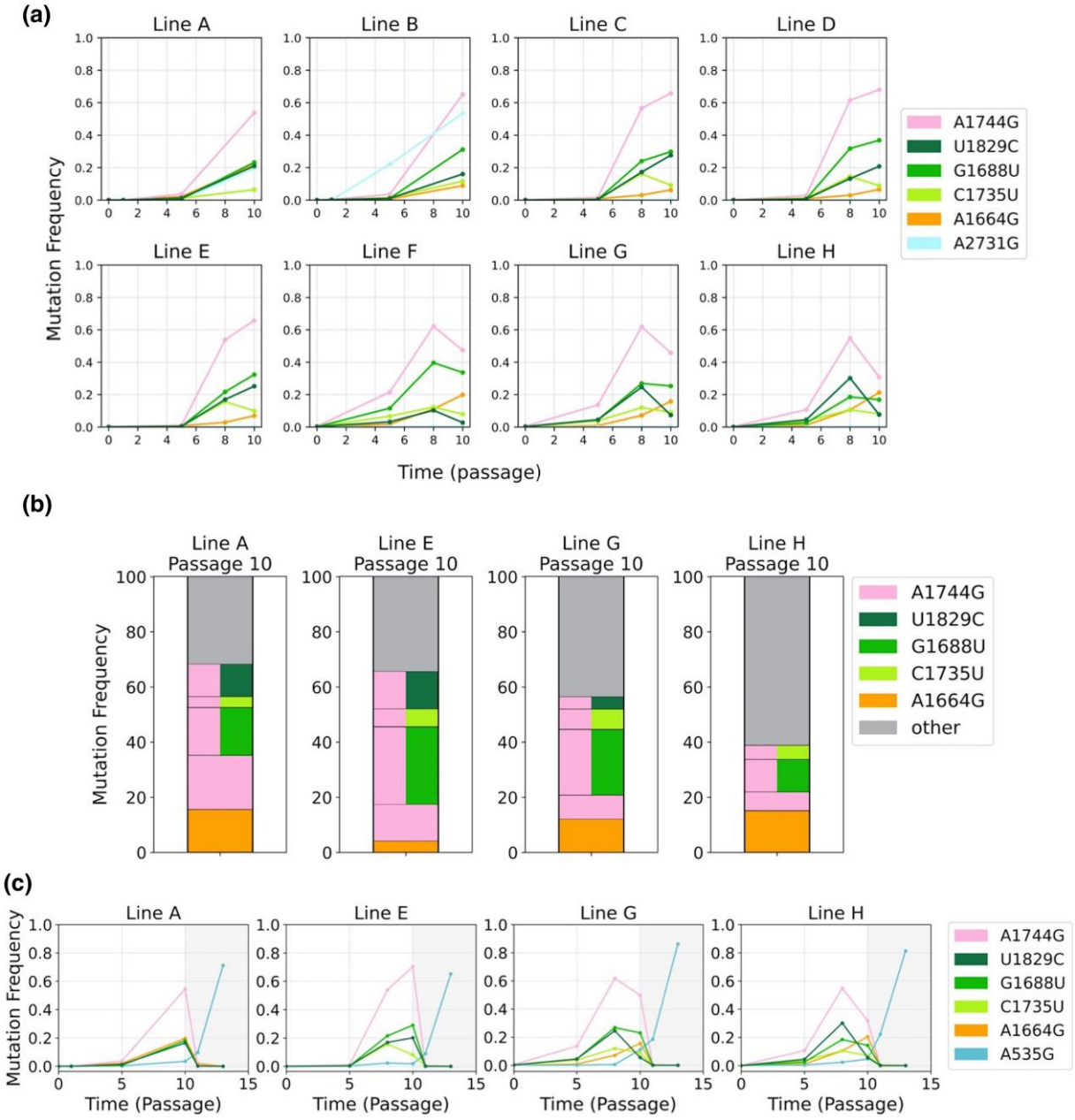


**Fig. 2.** Parallel evolution during MOI = 10 passaging. a) Trajectories of mutations from 8 replicas of serial passaging performed at an MOI of 10 reveal strong parallelism. Shown are mutations exceeding 10% in at least 2 or more replicas. b) Results of full haplotypes from 4 passage 10 populations (lines A, E, G, and H), derived from synthetic long-read sequencing. Each horizontal block represents a unique haplotype and is colored according to its mutations (colors coded as shown in a)). Block height is proportional to the frequency of the haplotype. Only haplotypes exceeding 3% are displayed to allow a view of mutations observed in a). c) Similar to a), with the addition of 3 terminal passages performed at a low MOI of 0.01. Passaging was reinitiated from a resuscitated passage 9 in 3 lines (A, E, G, and H), and 3 additional passages (11 to 13; gray background) were performed at an MOI of 0.01. Shown are mutations exceeding 10%. See supplementary fig. S5, Supplementary Material online for mutations at lower frequencies from a) and (c).

RNA structure that controls translation of the replicase protein, thus possibly reducing or enhancing replication, and (iii) alteration of the TR loop RNA structure, which halts replicase translation and initiates packaging, thus possibly reducing or enhancing packaging.

We thus set out to test the effects of A1744G-pink on replication using an assay of intracellular replication. We sequenced genomes from pellet (intracellular genomes) at 15, 30, 45, 60, and 90 min, as well as overnight (ON), using populations of p8-D. This was compared to supernatant (virion) sequencing at time points zero and ON. Two such assays were performed, 1 at an MOI of 0.01 (Fig. 3a) and 1 at an MOI of 10 (Fig. 3b). Importantly, the low MOI assay allowed us to infer how each of the possible genotypes behaves during singular infection, since such low MOI ensure each cell is infected by exactly 1

**Table 1** Inferred phenotypic effects of A1744G-pink and green secondary mutations G1668U, C1735U, and U1829C

| Mutation | Type | Affected protein/ structure | Inferred phenotype | Replication cycle stage (Fig. 1)[a,b] |
|---|---|---|---|---|
| A1744G | Nonsynonymous | Lysis (K23E) | Reduced lysis expression (Chamakura et al. 2017) | IV |
| | RNA structure | MJ | Alteration of the MJ RNA structure → reduction or enhancement of replicase translation (supplementary fig. S1, Supplementary Material online) | II |
| | RNA structure | TR | Alteration of the TR loop → reduction or enhancement of packaging (supplementary fig. S1, Supplementary Material online) | III |
| G1688U | Synonymous | Coat | – | – |
| | RNA structure | LH | Increased stability of RNA hairpin structure → decreased ribosome binding → reduced lysis expression (Berkhout et al. 1987; Schmidt et al. 1987) | IV |
| | Nonsynonymous | Lysis (R4L) | Unknown | IV |
| C1735U | Nonsynonymous | Lysis (R20W) | Reduced lysis expression (Chamakura et al. 2017) | IV |
| U1829C | Synonymous | Replicase | – | – |
| | Nonsynonymous | Lysis (F51S) | Reduced lysis activity (Chamakura et al. 2017) | IV |

[a]I—cell entry, II—replication, III—packaging, and IV—lysis.
[b]Replication cycle stages were inferred from the literature.

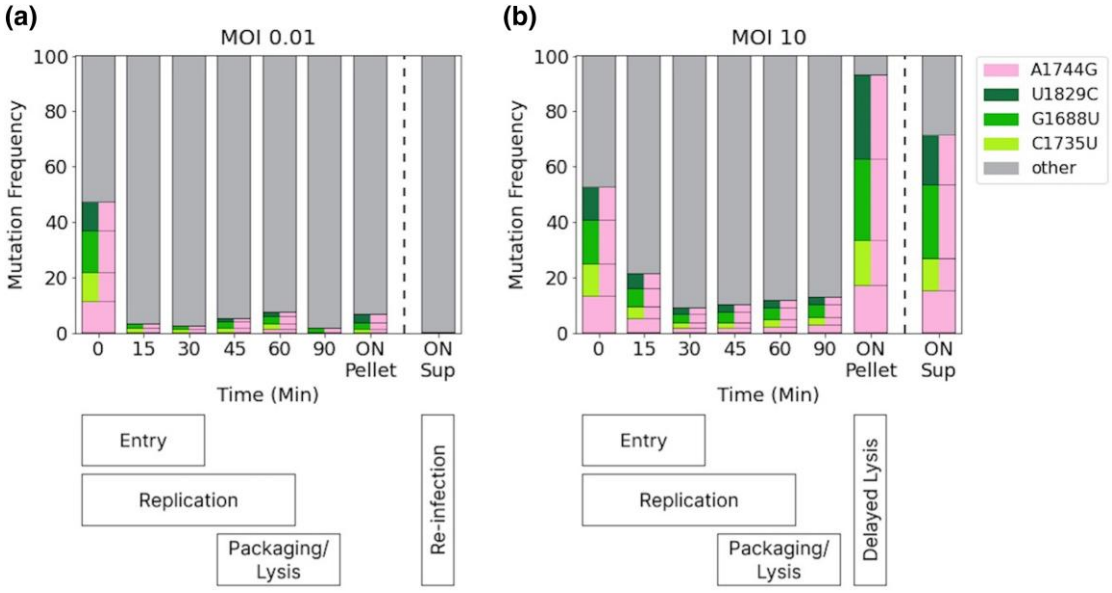


**Fig. 3.** Intracellular assay of replication for A1744G-pink and green secondary mutations. The frequencies of the various genotypes are shown across various time points of the viral replication cycle from genomes isolated from inside infected cells (pellet). Time points zero and ON-Sup represent sequencing of virions from supernatant at the start and end of a passage. Each horizontal block represents a unique haplotype and is colored according to its mutations (colors coded as shown in Fig. 2a). Block height is proportional to the frequency of the haplotype. Only genotype frequencies initially higher than 3% are displayed. Genotype frequencies were inferred from Fig. 2b (Methods). a) Results from the intracellular assay at MOI of 0.01, (b), Results from the intracellular assay at MOI of 10. The lower panels illustrate timing of entry, replication, and packaging/lysis based on plaque assay and qRT-PCR measurements (supplementary fig. S6, Supplementary Material online). Of note, replication is evident as of 7 min post infection (supplementary fig. S6, Supplementary Material online), but for simplicity is illustrated as starting at time zero. Presence of ~93% mutants in the ON Pellet sample at high MOI suggests delayed lysis, whereas increased presence (~71%) of mutants in ON Sup suggests increased packaging. Differences between low and high MOI timing stem from delayed entry at low MOI, as well as an additional replication cycle (reinfection) at low MOI.

genotype. Conversely, the high MOI intracellular assay allowed us to infer the behavior of genotypes under the conditions of the original serial passaging experiment.

The frequency of the combined A1744G-pink and green secondary haplotypes all dramatically declined between 0 and 15 min at both MOI 0.01 and 10. A 10-fold decrease was observed at MOI 0.01 (from ~52% to ~5%; $P < 10^{-10}$, Fisher exact test), and a 2- to 3-fold decrease was observed at MOI 10 (from ~52% to ~21%; $P < 10^{-10}$, Fisher exact test). These suggest either a replication disadvantage (stage II) or a cell entry disadvantage (stage I) associated with A1744G-pink. Since A1744G-pink was not inferred to impact entry in any way (Table 1), and neither were the green secondary mutations, we conclude that A1744G likely leads to a defect in replication due to its impact on the MJ structure, which causes a decrease in the

replicase protein production (Betancourt 2009). This defect is probably what led to no A1744G-pink in ON extracellular/sup populations at MOI of 0.01. The defect is compensated at higher MOI by coinfection and use of WT replicase, and hence less of a decline of A1774G-pink genomes is observed at high MOI at 15 min.

We noted a strong increase in frequencies of A1744G-pink and green secondary mutations in the ON pellet populations tested at MOI = 10 (7-fold increase, from ~13% at 90 min to ~93% ON; Fig. 3b; $P < 10^{-10}$, Fisher exact test). This result suggests either a replication advantage (which we ruled out above), or delayed lysis, allowing more genomes to remain intracellular. We see only minor frequencies (~7%) of A1744G-pink and secondary mutations in ON pellet populations at an MOI of 0.01, since there is likely another round/s of replication that releases WT viruses ON. Given that there is an inferred defect in lysis created by all of the green mutations (as well as for A1744G-pink) (Table 1) (Chamakura et al. 2017), we suggest that they all contribute to delayed lysis. Under conditions of high MOI (no more cells to re-infect) and under conditions of defective replication by A1744G-pink, delayed lysis may be an advantage as it provides additional time for genome replication of the cheaters.

What if so, is the advantage of A1744G-pink during high MOI that leads to the increase in A1744G-pink during serial passaging, also reflected in the ON sup populations at high MOI? The results suggest an advantage that must be post-replication. Given the potential effects of A1744G-pink on the TR loop structure (Table 1), we suggest that this advantage is in packaging. However, an alternative nonmutually exclusive explanation is that as described above, delayed lysis in itself provides an advantage under conditions where there is not an additional replication cycle: during the time the viruses are "trapped," or partially trapped, in the cells, replication continues. If lysis is not completely defective, some of these viruses may exit the cells during the long wait ON.

To summarize, we conclude that A1744G-pink is a replication-defective cheater that uses WT replicases as "public goods" during coinfection. A1744G-pink likely has an advantage in packaging and a possible defect in lysis as well. Green secondary mutations that emerge on the A1774G-pink background serve to delay lysis and allow more time for replication.

## Revisiting Serial Passaging at MOI of 1 Reveals More Secondary Mutations

In parallel to our serial passaging experiments at an MOI of 10, we had also previously performed serial passaging at MOI of 1 in 2 replicas, and originally, we observed competition between the 2 cheaters (A1664G-orange and Δ1764-red), with A1664G-orange emerging as a "winner" (Meir et al. 2020). We continued the passaging at an MOI of 1 till passage 30 (Fig. 4a), originally with the aim of testing (i) whether a steady state would be obtained and (ii) whether the semicheater A1664G-orange would "correct" its defect.

Our results showed that A1664G-orange continued to increase in frequency, attaining a frequency of over 90%. Once again performing long-read synthetic sequencing allowed us to resolve haplotypes (Fig. 4b). Interestingly, in both replicas we noted a temporary rise and demise of the A1744G-pink, occasionally with 1 of the green secondary mutations (G1688U) and occasionally with A1664G-orange. We also observed the rise and decline of A1664G-orange combined with A535G-turquoise (Fig. 4b). Finally, we observed a set of 3 so-called purple secondary mutations that began to increase in frequency as of passage 20, which all resided on genomes together with A1664G-orange. As opposed to the green secondary mutations described above, we did observe combinations of the purple mutations with each other. The decline of all combinations of mutations, versus the rise of combinations of A1664G-orange and the purple mutations, led us to conclude that the latter combinations are the most fit in the conditions of the passaging.

We noted that 2 of the 3 purple mutations (C1718U and C1549U) reside in the same region where there are multiple overlapping functions (Fig. 1b), rendering it challenging to infer their function (Table 2). The third, C2859U, creates a nonsynonymous mutation in the replicase gene, which resides 1 residue upstream a functional motif of the encoded protein (supplementary fig. S7, Supplementary Material online), lending a first clue as to the combined function of these mutations. In fact, in principle, all 3 mutations are inferred to impact replication (Table 2), as tested below.

## Does the High Frequency of A1664G-Orange Imply a Loss of Cheating Phenotype?

We considered it possible that the A1664G-orange semicheater "corrected" its defect via the secondary mutations, while retaining its advantage. To test this, we resuscitated the p30 samples and performed additional passaging at a low MOI of 0.01 that does not allow for coinfection (Fig. 4c). This assay revealed that the frequency of A1664G-orange and that of all the secondary mutations—crashed to zero. Two other mutations increased in frequency, A535-turquoise, described above, and U1691C, a mutation we have previously observed as adaptive under low MOI conditions probably via more rapid lysis of cells (Caspi et al. 2023). We conclude that the A1664G-orange semicheaters remain as cheaters during the entire duration of passaging in the MOI = 1 experiment.

## Mechanism of Purple Secondary Mutations

We set out to test whether we can infer the function of the purple secondary mutations. To this end, we once again performed an assay of intracellular replication, and sequenced genomes at 0, 15, 30, 45, 60, and 90 min, and ON, using populations of p28-B (Fig. 5). Importantly, as described above, MOI of 0.01 allows us to assume each cell is infected by 1 genotype, allowing us to make inferences about effects of each genotype on its own. MOI of 1 allows

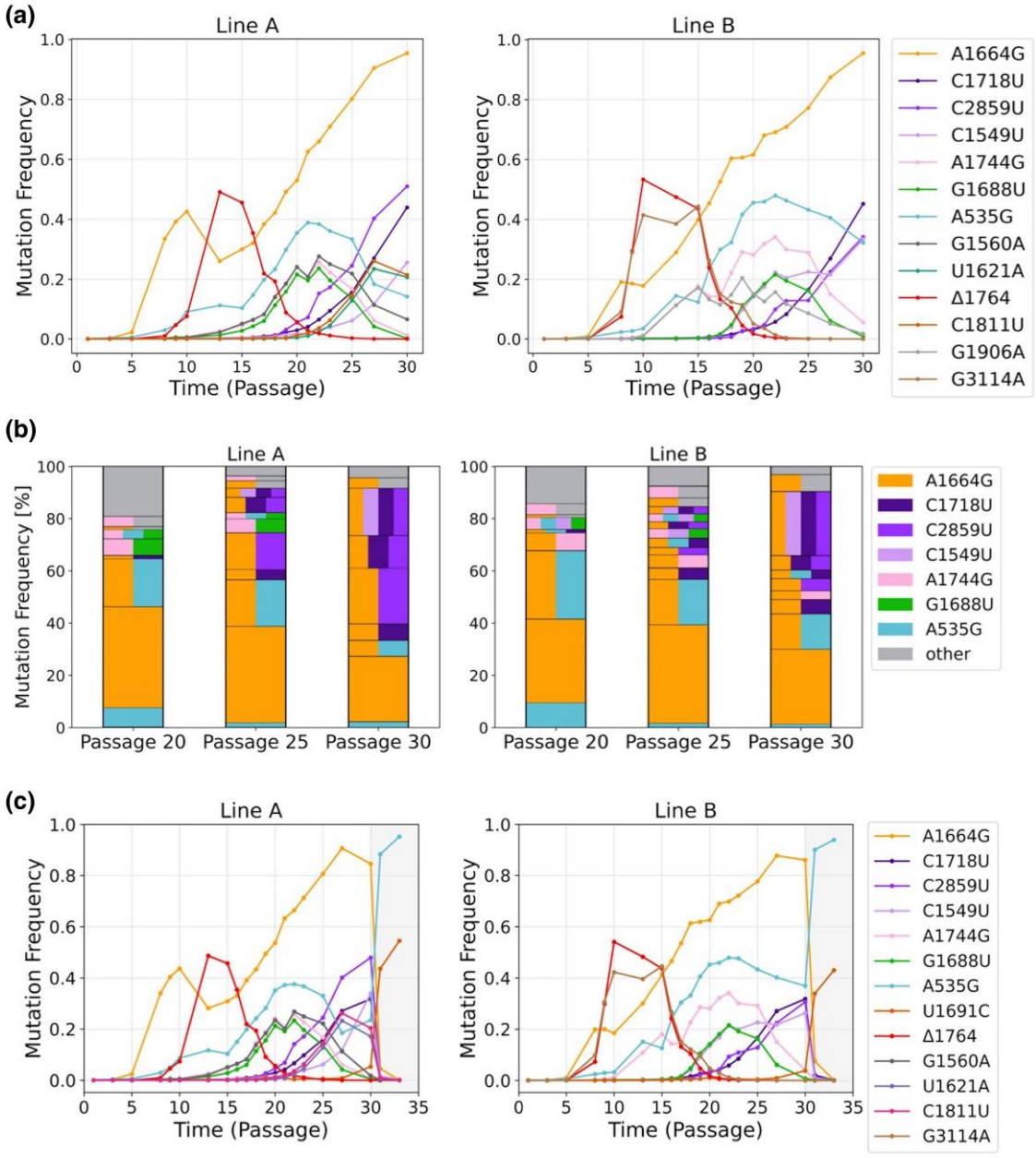


**Fig. 4.** Parallel evolution during advanced MOI = 1 passaging. a) Trajectories of mutations from 2 replicas of serial passaging performed at an MOI of 1, initiated in (Meir et al. 2020) and continued herein from a resuscitated p23 population. Due to the large number of mutations that emerge, shown are mutations with $f \geq 20\%$ in any replica. b) Results of full haplotypes from different passages from both lines. Each horizontal block represents a unique haplotype and is colored according to its mutations (colors coded as shown in a)). Block height is proportional to the frequency of the haplotype. Only haplotypes with $f \geq 3\%$ at one or more time points are displayed a). c) Similar to a), with the addition of 3 terminal passages performed at low MOI of 0.01. Passaging was reinitiated from a resuscitated p30 in both lines, and 3 additional passages (11 to 13; gray background) were performed at an MOI of 0.01. Shown are mutations with $f \geq 20\%$. See supplementary fig. S8, Supplementary Material online for mutations at lower frequencies from a) and c).

us to obtain a picture of what occurs during coinfection at the same conditions as the original experiment was performed.

First, as we have previously shown (Meir et al. 2020), A1664G-orange was not impacted in replication capacity, as is evident from the steady proportion of A1664G-orange during the entire first 60 min (~90%). During ON at MOI of 0.01, we observed a decrease in A1664G-orange in extracellular virions down to a frequency of ~30%, reiterating its ability to replicate on its own that is inferior to WT (semi-cheater phenotype) (Fig. 5a).

When observing A1664G-orange in the MOI = 1 intracellular assay (Fig. 5b), we saw a more or less steady proportion of A1664G-orange across all time points (~90%), with a slight increase in intracellular ON genomes (~96%). The latter reiterated the defect in lysis we have previously shown (Meir et al. 2020). In general, the population used here for the intracellular replication

**Table 2** Inferred phenotypic effects of A1664G-orange and purple secondary mutations C1549U, C1718U, and C2859U

| Mutation | Type | Affected protein/ structure | Inferred phenotype | Replication cycle stage (Fig. 1)[a,b] |
|---|---|---|---|---|
| A1664G | Synonymous | Coat | – | – |
| | RNA structure | LH | Reduced lysis translation (Meir et al. 2020). | IV |
| C1549U | Nonsynonymous | Coat (T72I) | Reduced coat protein binding to packaged RNA genome, may alter virion structure and/or alter RNA release during cell entry. | I |
| | | | Reduced coat protein binding to TR loop, may delay packaging and increase replication (Stonehouse et al. 1996; Lago et al. 1998). | II/III |
| C1718U | Synonymous | Coat | – | – |
| | RNA structure | CT | Destabilizes CT (supplementary fig. S1, Supplementary Material online) → may increase replicase translation and hence replication (Van Himbergen et al. 1993). | II |
| | Nonsynonymous | Lysis (A14 V) | Unknown | IV |
| C2859U | Nonsynonymous | Replicase (R367C) | Adjacent to predicted replicase motif E (RNA binding) (supplementary fig. S7, Supplementary Material online). May promote or decrease replication. | II |

[a]I—cell entry, II—replication, III—packaging, and IV—lysis.
[b]Replication cycle stages were inferred from the literature.

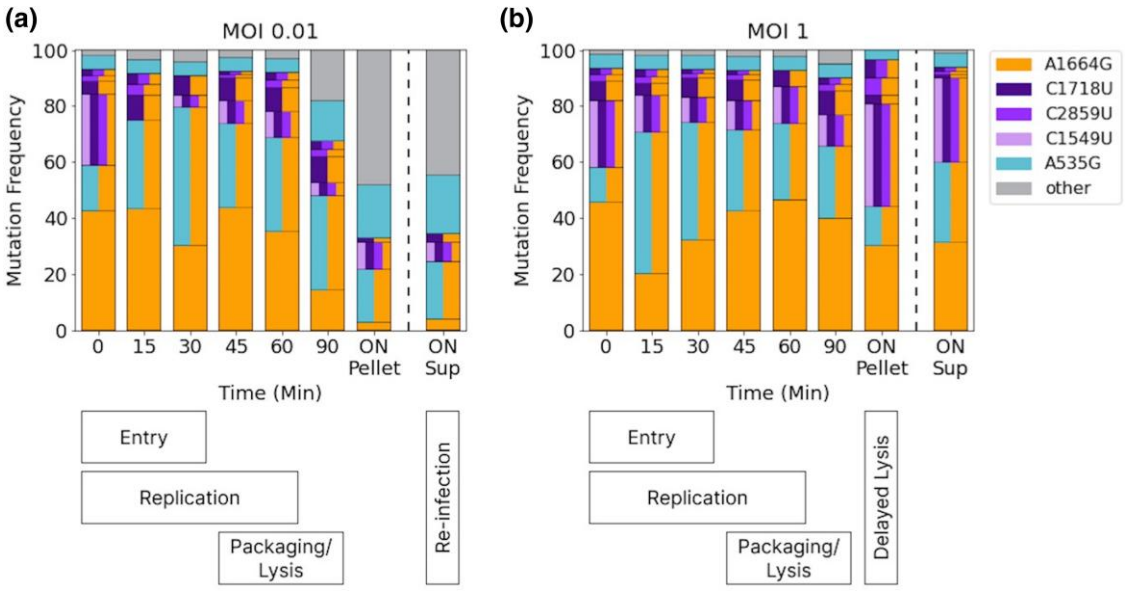


**Fig. 5.** Intracellular assay of replication for A1664G-orange and additional mutations. The frequencies of the various genotypes are shown across various time point of the viral replication cycle from genomes isolated from inside infected cells (pellet). Time points zero and ON-Sup represent sequencing of virions from supernatant at the start and end of a passage. Each horizontal block represents a unique haplotype and is colored according to its mutations (colors coded as shown in Fig. 4a). Block height is proportional to the frequency of the haplotype. Only genotype frequencies initially higher than 5% are displayed. Genotype frequencies were inferred from Fig. 4b (Methods). a) Results from intracellular assay at MOI of 0.01, and b) Results from intracellular assay at MOI of 1. The lower panel illustrates timing of entry, replication, and packaging/lysis (see Fig. 3 legend).

assay was much more complex (many more mutations and many more combinations) than the population used for the intracellular assay described earlier for A1744G-pink (Fig. 3). This necessitated us to develop a simple model of intracellular replication at low MOI, with fitness parameters $w$ for 3 of the 4 replication cycle stages where we infer an impact by the various mutations: entry, replication, and lysis. Lysis effectively captures the 2 last stages together of packaging and exit. An additional nuisance parameter was added to allow for reinfection during low MOI (Methods).

Each parameter was assumed to affect the frequency of genotypes at a specific stage of intracellular replication. Thus, for example, a mutant with more rapid replication than WT, would be associated with a parameter $w_r > 1$ that allows for an increase in the frequency of this mutant at times points where replication occurs (15 through 60 min; Methods, supplementary fig. S6, Supplementary Material online). The model allowed us to simulate the trajectory of intracellular frequencies of a mutant given its associated fitness values, and to fit this trajectory to the observations depicted in Fig. 5. We next used an

approximate Bayesian computation (ABC) approach to infer parameters for the genotypes shown in Fig. 5.

To avoid over-parameterization, we combined all purple mutations together and referred to them as the "purple combo." Our inferences suggested that the purple combo negatively impacted cell entry (posterior probability > 0.99), but also had an advantage during intracellular replication (posterior probability > 0.98). We did not find evidence to support or rule out that the purple combo led to more delayed lysis (supplementary fig. S9, Supplementary Material online). These former results can be derived by observing the MOI = 1 intracellular assay results: we saw a decrease in the purple combo from 0 to 15 min, supporting a cell entry defect, and an increase in purple combo over time that was particular evident in the ON pellet sample, supporting an advantage in replication.

To summarize, we conclude that A1664G-orange is a lysis-defective cheater that uses WT lysis as "public goods" during coinfection. A1664G-orange likely has an advantage in packaging as ribosomes not translating lysis will create more coat protein. Purple secondary mutations that emerge on the A1664G-orange background serve to increase the replication rate thus allowing for more lysis to be translated. The latter advantage becomes critical as the frequency of A1664G-orange increases and the probability of coinfection with WT decreases.

## Discussion

In this work, we set out to discover the repertoire of cheaters in the MS2 bacteriophage, and their subsequent evolution during high MOI conditions. We detected 3 point-mutant cheater phages, each of which possessed a dual characteristic: defective at low MOI and advantageous over WT during high MOI. Interestingly, as opposed to work in previous viruses, we did not note any structural variants and in particular did not note high frequencies of large deletions at high MOI. It is possible that large deletion cheaters exist in MS2, but that they are outcompeted by the potent cheating of the point-mutant cheater viruses. Moreover, as packaging signals exist throughout the MS2 genome (Rolfsson et al. 2016; Dai et al. 2017), this may render large deletion cheating less effective.

It was intriguing to note that all the 3 cheaters we have found here and previously (Δ1764-red, A1664G-orange, and A1744G-pink), and most of the so-called secondary mutations that subsequently evolved on the background of 2 of these cheaters, resided in a region spanning <300 nucleotides (Fig. 6a), i.e. <10% of the genome. This region is information-rich, and many mutations are expected to have pleotropic effects: there are 2 overlapping ORFs, and a series of RNA structures that overlap 1 or 2 of these ORFs. Analyzing the evolutionary rate of this region among homologs of MS2, we find that this region is one of the more conserved regions, displaying for most part a low rate (supplementary fig. S10, Supplementary Material online). Yet, we find that cheaters specifically tend to evolve in this region. We suggest that this stems from the nature of point-mutant cheaters—they must both disrupt 1 function and gain an advantage of another function, and this is possible exactly at regions of overlapping functions. Cheaters "escape" the rigidity of selection operating at these regions due to their reliance on coinfection.

We go on to discuss the role of secondary mutations in modulating the effects of the A1744G-pink mutation. In this case, the 3 additional green secondary mutations were all inferred to be deleterious to lysis production or function. As proposed, delaying lysis can be advantageous when (i) there are no additional hosts to infect at the end of each infection cycle (high MOI), and (ii) replication is impaired on the background of A1744G-pink, and delayed lysis grants more time for replication (Fig. 6b). This raises questions about why the specific green mutations are involved in lysis delay and why they do not appear together on the same genome. The MS2 lysis protein remains enigmatic, and recent studies suggest it functions as an oligomeric complex (Mezhyrova et al. 2023). We propose that lysis delay arises from mutants that partially "poison" these oligomeric complexes, composed of WT and mutants during coinfection. A similar phenomenon was recently observed in influenza polymerase units, where a subunit produced by DVGs was found to impair the activity of the entire polymerase complex (Ranum et al. 2024). Consequently, double lysis mutants or other lysis mutants (beyond the 3 identified) may induce a dominant-lethal phenotype, resulting in either rapid lysis or complete lysis inhibition.

We noted that in some of the replicas of the MOI = 10 serial passaging, there was a decline of A1774G-pink haplotypes that was concurrent with an increase in the frequency of the A1664G-orange haplotype. In fact, we previously observed a similar ecological interaction between Δ1764-red and A1664G-orange (Meir et al. 2020). This suggests that A1664G-orange outcompetes A1774G-pink; yet, this possible interaction between the 2 cheaters requires further research.

Moving on to the advantage of the purple secondary mutations on the A1664G-orange background, we noted the dynamics became more intricate. One possibility is that all 3 purple mutations enhance replication (Fig. 6c), which would be beneficial under the delayed lysis conditions induced by A1664G-orange; increased genome replication could lead to higher lysis protein production, compensating for the delay and enabling cell lysis. This scenario is, in some respects, a mirror image of the effects observed with A1744G-pink and its green secondary mutations. Specifically, for C2859U-purple, replication likely is impacted since this mutation introduces a nonsynonymous change in the replicase protein. C1549U-purple and C1718U-purple may also delay replication, though they could produce additional effects independently. C1718U-purple might further influence lysis or replication through an unknown mechanism (Table 2), while C1549U-purple appears to delay cell entry—a disadvantage potentially offset by its capacity to enhance replication. Notably, this pattern suggests that when a cheater

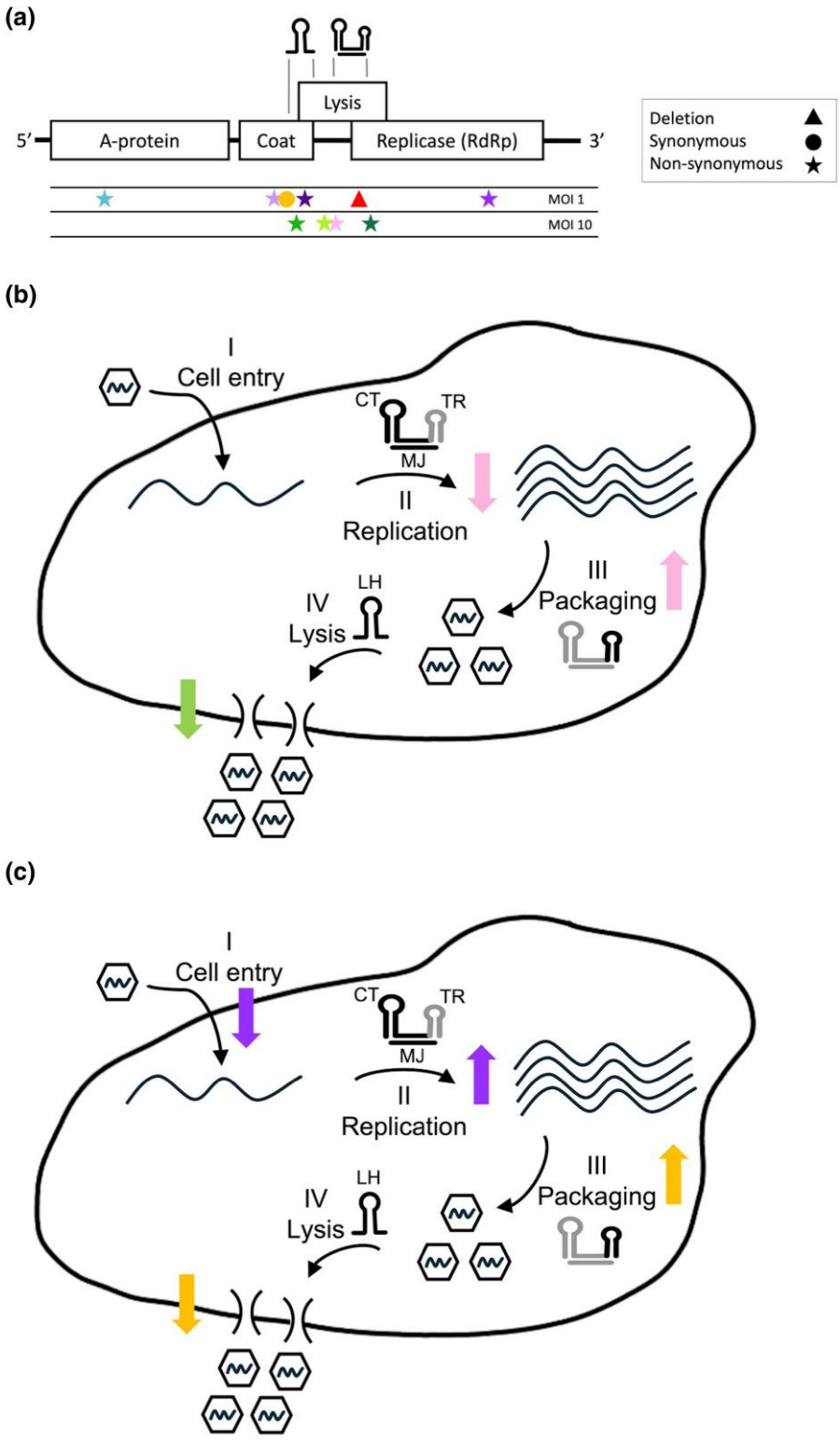


**Fig. 6.** Illustration of viral genome with accumulated mutations and fine balance model of viral replication cycle during viral cheating. a) A schematic of the MS2 genome. Below the genome is a summary of all the major mutations found herein and previously in the context of high MOI serial passaging. Most mutations observed in this work resided in a region spanning <300 nucleotides. b and c) Shown is the disruption of the cycle by cheaters and correction by secondary mutations for the scenario of the A1744G-pink cheater b) and the A1664G-orange semi-cheater c). The 4 main stage of viral replication are shown as in Fig. 1a. Upwards or downwards colored arrows represent an increase or decrease in rates of a relevant stage based on the inferences deduced herein. The illustration describes a singular infection by a given genotype, and the number of phages that exits the cell is not modelled in the illustration.

mutation disrupts 1 or more stages of the replication cycle, subsequent mutations on this background modulate other stages to restore balance. This points to a delicate equilibrium in the viral replication cycle, where "pulling the strings" of different stages can shift the cycle's dynamics.

We compare between the primary and secondary cheater mutations we found herein for MS2, to findings in other viruses. Overall, point-mutant cheater viruses have been rarely reported (Fodor et al. 2003; Dennehy and Turner 2004). They may be overlooked in many other RNA viruses, as they are often difficult to infer, and we suggest regions with overlapping functions may be a good starting point for such searches. On the other hand, cheaters/DVGs based on large deletions or copy-back were detected across most classes of RNA viruses and were found both in vitro and in vivo (Enea and Zinder 1975; Brennan and Sun 2024).

To the best of our knowledge, secondary mutations have rarely been reported for DVGs. A notable variety of a secondary mutation was observed in f1 phage, where secondary mutations tend to occur on the background of the WT phage during coinfection with the cheater (Horiuchi 1983), rather than on the background of the cheater as we found herein. Overall, it is intriguing to consider that cheaters may also serve as stepping stones in evolution, which allow viruses to overcome the stringent selection operating in such overlapping function regions. In other words, cheaters may evolve into cooperators, a concept that has been widely discussed under the umbrella of multilevel selection (Fiegna et al. 2006; Traulsen and Nowak 2006; Simon et al. 2013).

This leads to the discussion on how prevalent high MOI conditions are in a natural setup. First, it has been shown that several viruses transit via collective infectious units (Sanjuán 2017; Leeks et al. 2019), which is somewhat akin to higher MOI as it allows multiple virus genotypes to enter cells. More generally, MOI probably tends to cycle between high and low in most conditions, similar to predator prey cycles, and thus cheaters probably emerge and disappear cyclically. However, we believe that high MOI may possibly remain for a few replication cycles, as is observed during within-host evolution of many animal viruses (Vignuzzi and López 2019; Felt et al. 2021). Over billions of years of evolution, cheaters will thus constantly emerge and evolve, potentially facilitating exploration of the fitness landscape in ways that could result in significant genetic innovations.

## Materials and Methods

### Stock Preparation of MS2 Phage

MS2 phage (from ATCC 15597-B1) and the F pilus expressing donor *Escherichia coli* c-3000 were cultured in LB medium at 37 °C with shaking ON. Bacteria cells were removed by centrifugation at 4,000 rpm for 20 min at 25 °C. The supernatant was filtered with a 0.22 μm filter (Stericup® Filter, EMD Millipore) to remove any remaining residues. The phage was then stored at 4 °C. Typically, a stock concentration of $10^{11}$-$10^{12}$ plaque forming units (PFU)/ml was obtained.

### MS2 Plaques Isolated

All the experiments started using a single WT MS2 plaque collected during plaque assay. A single MS2 plaque was collected with a sterile Pasteur pipette and transferred to an Eppendorf tube containing 1 ml of 0.85% NaCl. The tube was stored at 4 °C ON to facilitate viral particle release from the agarose. Next, 10 μl of chloroform was added, followed by vortexing and centrifugation for 3 min at 13,000 rpm at room temperature. The supernatant was then filtered through a 0.22 μm Minisart syringe filter (Sartorius) to remove residues. RNA was extracted from each plaque for whole genome deep sequencing, as detailed below. In addition, plaques in 3 different sizes (small, medium, and large) were isolated from p8-D stock using the same method.

### Serial Passaging

Clonal MS2 stocks were propagated from single plaques to ensure that the experiments will begin with a phage population as homogeneous as possible. Each clone served as the precursor for all evolutionary lines in each experiment. Serial passage experiments were conducted with 2 or 3 biological replicates, as followed. Three similar experiments were performed under the same conditions of MOI 10, Lines A to H. The first experiment included lines A and B, the second experiment included lines C to E and the third experiment included lines F to H. An additional experiment previously described (Meir et al. 2020) of serial passaging at MOI 1 was performed in 2 biological replicates, and then denoted as lines 37A and 37B. The procedure for serial passaging was as follows: 100 ml cultures of naive *E. coli* c-3000 were grown at 37 °C to an optical density of $OD_{600} = 0.5$, corresponding to $\sim 10^8$ bacterial cells/ml (total of $10^{10}$ bacterial cells). For a MOI of 10, each passage was infected with 10 ml of $10^{10}$ phages (PFU) from the previous passage (total of $10^{11}$ PFU); for an MOI of 1, each passage was infected with 10 ml of $10^9$ phages from the previous passage (total of $10^{10}$ PFU), maintaining an MOI of about 10 and 1 PFU/cell, respectively, at the start of each passage. The cultures were incubated at 37 °C for ON with shaking, and the *E. coli* cells were subsequently removed by centrifugation. The supernatant was filtered through a 0.22 μm filter (Stericup® Filter, EMD Millipore) to eliminate any remaining residues. The resulting phage stocks were then stored at 4 °C. Aliquots of these stocks were used to measure phage concentration via plaque assays, infect subsequent serial passages, isolate RNA for whole genome deep sequencing (as detailed below), and maintain frozen stocks of the evolving lines in 15% glycerol at −80 °C.

For the experiments testing the effects of lowering the MOI, additional low MOI passages were conducted using the same protocol. This time, a frozen stock of either p30 line 37A MOI 1 and p30 line 37B MOI 1 or p10-C and p10-D was used as the starting point to test the effect of a lower MOI of 0.01 on the cheater population.

### Plaque Assay

A plaque assay procedure was used to determine the concentration of the MS2 phages at each passage. The plaque assay was conducted according to the method described by Meir et al. (2020).

### RNA Isolation

RNA was isolated using 2 methods based on the experiment:

1) Typically, RNA from all serial passages stocks as well as RNA from the supernatant of the replication assay was isolated using the QIAamp® viral RNA mini kit (Qiagen) according to the manufacturer's instructions.

2) For the purpose of replication assays that test intracellular RNA, total RNA from *E. coli* cells (pellet) was isolated using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions.

### Illumina NGS Library Preparation

The MS2 RNA was reverse transcribed using SuperScript® 4 Reverse Transcriptase (Thermo Scientific), using the 3-R3 primer (supplementary table S1, Supplementary Material online). cDNA from the reverse transcription reaction was directly used as a template for the PCR amplification of the full MS2 genome in either 3 or 7 overlapping fragments, and PCR reactions were performed using the Phusion High-Fidelity DNA-polymerase (Thermo Scientific) according to manufacturer instructions using the primers that are listed in supplementary table S1, Supplementary Material online. As quality control, we sequenced a set of given samples using either 3 or 7 amplicons and the sequencing results were reproducible (not shown). The PCR products were purified using Wizard Gel and PCR Cleanup System (Promega). Purified amplicons were diluted and pooled in equimolar concentrations. To produce DNA libraries, the Illumina Nextera XT library preparation protocol and kit (FC-131-1096) were used according to manufacturer instructions. Libraries were sequenced on an Illumina Miseq or Nextseq using various reagent kits for paired end reads (MS-103-1002 Illumina, MS-102-2003 Illumina). All libraries were prepared in house and sequenced in the Genomics Research Unit at Tel-Aviv University (Israel), with the exception of the intracellular assay libraries that were prepared and sequenced by the Technion Genomics Center, Technion-Israel Institute of Technology, Haifa, (Israel).

### Read Alignment and Analysis of Illumina NGS

Illumina Miseq fastq files were initially processed using adapter removal and trimming. Adapter removal was performed using the Trimmomatic tool (Bolger et al. 2014) using ILLUMINACLIP for paired end sequencing with parameters: 2 seed mismatches, 30 palindrome clip threshold, and 10 simple clip threshold. Trimming was performed with seqtk tool (https://github.com/lh3/seqtk) using trimfq to cut 30 bases from the beginning and end of each read. Analysis of the sequencing data was performed using the AccuNGS pipeline (Gelbart et al. 2020) with the parameters: minimal %ID = 85, e-value threshold = 1E-07 and q score cutoff of 30. The pipeline uses BLAST (Altschul et al. 1990; McGinnis and Madden 2004) to map the reads to the reference genome, searches for variants that appear on both overlapping reads and calls variants with a given Q-score threshold. Finally, it infers the frequency of each variant in the population. The reference genome is as described by Meir et al. (2020). Libraries attained a minimum coverage of 100 reads per base. Positions at both ends of the genome (at the primer positions plus 5 adjacent positions) were removed from analysis because they had low coverage and higher variability.

### Library Construction, Sequencing, and Processing of Synthetic Long-Reads

RNA samples from passages p20, p25, and p30 from both lines (37A and 37B) at MOI 1 were sent to Element Biosciences (San Diego, CA, USA) for sequencing. The sequencing was performed using the LoopSeq RNA preparation kit following the manufacturer's instructions (information available at elementbiosciences.com). The LoopSeq protocol (Callahan et al. 2021) utilizes unique molecular barcoding technology, which distributes barcodes evenly across a genome before fragmenting it into shorter pieces. These labeled fragments are then sequenced using short-read sequencing methods on the AVITI platforms, followed by the reconstruction of full-length genomes. The short-read raw data were processed through the Loop Genomics analytical pipeline, which handles low-quality base trimming, unique sample barcode demultiplexing, and synthetic long-read reconstruction. This process allows for de novo assembly of full-length genomes by rearranging short reads tagged with the same unique barcode.

We used BLAST (Altschul et al. 1990; McGinnis and Madden 2004) to align the long-reads obtained from LoopSeq with the MS2 reference sequence (Meir et al. 2020). The alignment was performed using the following parameters: –evalue 1e-07–perc_identity 0.85–task blastn–num_alignments 1,000,000–dust no–soft_masking F. To ensure that only reads spanning the full alignment were considered, we filtered out any alignments that mapped more than once to the reference and those shorter than 3,500 nucleotides (98% of the reference genome length). Additionally, we excluded alignments that aligned to the minus strand, i.e. those that were the reverse complement of the genome.

### Intracellular Replication Assay

*E. coli* c-3000 cells were grown to an optical density of $OD_{600} = 0.5$ at 37 °C. 100 ml of bacteria were infected with 10 ml of different viral populations: p28 line 37B at MOI 1 and p8-E, each one diluted to produce an MOI of 0.01, 1 or 10 PFU/cfu. The cultures were grown at 37 °C with shaking. 1 ml volumes of the cultures were removed at 0, 15, 30, 45, 60, and 90 min post infection, and after growth ON. Infection was stopped and the bacteria were separated from the supernatant by centrifugation at 13,000 rpm for 1 min at 4 °C. The cells were lysed by resuspension in 0.2 mg lysozyme and incubated at room temp for 10 min. Viral RNA, from both cells and supernatant, was extracted as described above (MS2 RNA isolation). Illumina NGS library was prepared and analyzed as described above.

### Haplotype Inference for Intracellular Replication Assay

To generate the replication assay plot, we had to infer haplotype data despite the fact that the populations were sequenced using short-read Illumina sequencing.

**Table 3** Parameters $\omega_{s,m}$ of the intracellular replication model

| | Orange + purple combo | Orange + turquoise | Turquoise | Orange | wt |
|---|---|---|---|---|---|
| **Entry** | $\omega_{e,oc}$ | $\omega_{e,tu}$ | $\omega_{e,tu}$ | 1 | 1 |
| **replication** | $\omega_{r,oc}$ | 1 | 1 | 1 | 1 |
| **lysis** | $\omega_{l,o\ +\ \delta}$ | $\omega_{l,o}$ | 1 | $\omega_{l,o}$ | 1 |
| **re-infection** | $\omega_{rei,o}$ | $\omega_{rei,o}$ | $\omega_{rei,tu}$ | $\omega_{rei,o}$ | 1 |

However, we were able to rely on LoopSeq synthetic long-reads from previous experiments (Figs. 2b and 4b). In the MOI 10 experiment, each of the green secondary mutations was never found one with each other, and almost always appeared in combination with the A1744G-pink mutation, allowing us to assume the same behavior during the replication assay. In the MOI 1 experiment, we made the following assumptions: (i) Mutation C1549U (light purple) in line 37B only appeared as part of a quadruplet with the other purple mutations and A1664G-orange, so we used its frequency for the entire group and reduced the frequencies of individual mutations accordingly. (ii) Mutation C2859U (purple) appeared either as a pair with A1664G-orange or as a triplet with C1718U (dark purple) and A1664G-orange at similar frequencies; we divided C2859U's remaining frequency equally between the 2. We then adjusted C1718U and A1664G-orange frequencies by subtracting the triplet and pair frequencies, respectively. (iii) The remaining frequency of C1718U was assumed to be in combination with A1664G-orange. (iv) At MOI 1, mutation A535G-turquoise was present at no more than 5%, with the remainder assumed to occur on the same haplotype as A1664G-orange. At MOI 0.01, from 0 to 60 min, the same haplotype structure as in MOI 1 was assumed. At 90 min, the pair of A1664G-orange and A535G-turquoise remained at the same frequency as at 60 min. At the ON timepoints, A535G-turquoise was considered to be 50% on its own and 50% in combination with A1664G-orange. This approach allowed us to piece together short Illumina reads into coherent haplotypes, ensuring that the replication assay accurately reflected the underlying genetic structure. Notably, there was very low coverage of amplicon 1 (first third of the genome) in sample 15 min MOI 0.01, in the MOI 1 experiment. We re-sequenced this amplicon to reliably infer the frequency of mutation A535G-turquoise reported above.

## Intracellular Replication Model

We present a mathematical model that describes the intracellular replication cycle, encompassing the stages of entry, replication, and lysis. The lysis stage is assumed to capture both the packaging and exit processes. Additionally, we introduce a reinfection stage to account for reinfection at low MOI. Each parameter in the model (Table 3) is assumed to affect the frequency of specific genotypes at various stages of intracellular replication. Thus, for example, more rapid replication of a mutant would lead to increased frequency of this mutant at times points where replication occurs (15 through 60 min; Methods, Fig. 1). We next used an ABC (Beaumont et al. 2002) approach to infer parameters for specific genotypes.

Let $\omega_{stage,\ mutation} \in \mathbb{R}^+$ be the parameters of our model, where the stage is denoted as $s \in \{entry\ (e),\ replication\ (r),\ lysis\ (l),\ re-infection\ (rei)\}$ and the mutations (haplotypes) as $m \in \{A1664G,\ A1664G\ and\ purple\ helper\ mutations,\ A535G, A1\ 664G\ and\ A535G,\ wt\}$.

We define: $o\ (orange) := A1664G,\ oc\ (orange + combo) := A1664G\ and\ purple\ combo\ helper\ mutations,\ tu\ (turquoise) := A535G, otu\ (orange + turquoise) :=,\ A1664G\ and\ A535G.$

Hence, $m \in \{o,\ oc,\ tu,\ otu,\ wt\}$.

Each parameter $\omega_{s,m}$ represents the "fitness" of the haplotypes $m$ at the stage $s$ of the replication cycle. We also introduce $\delta \in \mathbb{R}$ as the parameter representing the added delay in lysis due to the presence of the combination of the purple secondary mutations over the orange mutation.

To avoid over-parameterization, we made the following assumptions:

1) Parameters for haplotypes that do not affect certain stages of the replication cycle were fixed at 1. Thus, for example, a nonsynonymous mutation in the lysis protein was assumed to have no impact on entry. Accordingly, entry parameters were assumed only for nonsynonymous mutations at A or coat proteins. Replication parameters were assumed only for nonsynonymous mutations at the replicase protein, mutations that affected the MJ structure or the CT structure, or mutations affecting the coat protein that interacts with the TR loop. Lysis parameters were assumed for nonsynonymous mutations at the lysis protein or mutations affecting the LH structure.
2) For haplotypes consisting of multiple mutations, the parameters were determined by the mutation most relevant to the specific stage of the replication cycle. For example, the orange mutation (A1664G) affects the lysis, therefore $\omega_{l,otu} = \omega_{l,o}$.

## Simulation of Haplotype Frequencies

We simulated the frequencies of the observed haplotypes at each time point in our second intracellular replication assay (based on the original passaging experiment at MOI = 1), which was conducted at MOI of 0.01 (Fig. 5a). The objective was to infer the effects of individual mutations or combinations of mutations on the different stages of the replication cycle.

The set of time points is defined as $T = \{t_0, t_{15}, t_{30}, t_{45}, t_{60}, t_{75}, t_{90}\}$, where frequencies are simulated at each time point. Notably, $t_{75}$ is simulated solely

as a correction for uniform time intervals of 15 min, with no data fitting at that time point, as no empirical data exists for it.

Based on supplementary fig. S6, Supplementary Material online, we define indicator vectors $I_t \in \{0, 1\}^4$, which indicate the active stages—entry, replication, lysis, and reinfection respectively, at each time point, as follows:

$$I_{t_{15}} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \; I_{t_{30}} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \; I_{t_{45}} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, I_{t_{60}} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix},$$

$$I_{t_{75}} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}, I_{t_{90}} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

Uncorrected haplotype frequencies at a given time $t$ are given by:

$$P_t^m = P_{t-1}^m \cdot \prod_{s \in S} \omega_{s,m}{}^{I_t}$$

To compute the corrected frequencies, we normalize the haplotype frequencies by calculating the relative abundance:

$$f_t^m = \frac{P_t^m}{\sum_{m \in M} P_t^m}$$

Initial empirical frequencies at time point zero $t_0$ were used to initialize the simulations.

### Inference of Model Parameters

We employed a wide prior for the "fitness" parameters thus minimizing prior assumptions and allowing for a broad range of parameter sampling, with $\omega_{s,m} \sim U(0, 4)$ *for all* $s \in$ *Stages and* $m \in$ *Mutations*. For the additional lysis delay parameter, representing the possible effect of the purple secondary mutations (denoted as combo), we used a prior $\delta \sim U(-1, 1)$, aiming to assess whether these secondary mutations contribute additional delay beyond the orange mutation (A1664G) alone.

Given the complexity and large number of parameters, traditional maximum likelihood estimation was impractical. Therefore, we utilized ABC with sequential Monte Carlo (ABC-SMC) (Lintusaari et al. 2017), as implemented in the Python package pyABC (Klinger et al. 2018). As a summary statistic, we used the sum of squared differences between the simulated and empirical data across all time points, calculated using the following squared $\ell_2$ distance function:

$$\ell_2^2(f_{simulated}, f_{empirical}) = \sum_{t \in T} \sum_{s \in S} (f^s_{simulated_t} - f^s_{empirical_t})^2$$

The ABC-SMC algorithm was run with threshold $\varepsilon$ that decreased to 0.078 over 15 iterations. Each iteration used 10,000 particles, and the particles of the final iteration were used to generate posterior distributions for all parameters.

## Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

## Acknowledgments

We thank Avigdor Eldar and Danielle Miller for stimulating conversations and insightful reading. This study was supported by an ERC starting grant 852223 (RNAVirFitness) to A.S. and by a European Research Council grant AdG 787514 to U.G. This study was also supported by a fellowship to Y.M., N.H., and A.B.Z. from the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University.

## Data and Code Availability

All sequencing data presented in this paper are available in the sequencing read archive (SRA). Previously published data (MOI = 1 passages 1 through 23) are available under BioProjects PRJNA575138, PRJNA547685. All other sequencing data are available under BioProject PRJNA1161593. Processed data including inferred mutation frequencies and haplotypes are available in the Zenodo database under accession code 10.5281/zenodo.14196524. All code used for sequence analysis and for modelling and parameter inference is available at https://github.com/Stern-Lab/cheaters_fine_balance.

## References

Altan-Bonnet N, Chen Y-H. Intercellular transmission of viral populations with vesicles. J Virol. 2015:**89**(24):12242–12244. https://doi.org/10.1128/JVI.01452-15.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990:**215**(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Bdeir N, Arora P, Gärtner S, Hoffmann M, Reichl U, Pöhlmann S, Winkler M. A system for production of defective interfering particles in the absence of infectious influenza A virus. PLoS One. 2019:**14**(3):e0212757. https://doi.org/10.1371/journal.pone.0212757.

Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. Genetics. 2002:**162**(4):2025–2035. https://doi.org/10.1093/genetics/162.4.2025.

Berkhout B, Schmidt BF, van Strien A, van Boom J, van Westrenen J, van Duin J. Lysis gene of bacteriophage MS2 is activated by translation termination at the overlapping coat gene. J Mol Biol. 1987:**195**(3):517–524. https://doi.org/10.1016/0022-2836(87)90180-X.

Betancourt AJ. Genomewide patterns of substitution in adaptively evolving populations of the RNA bacteriophage MS2. Genetics. 2009:**181**(4):1535–1544. https://doi.org/10.1534/genetics.107.085837.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014:**30**(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brennan JW, Sun Y. Defective viral genomes: advances in understanding their generation, function, and impact on infection outcomes. Mbio. 2024:**15**(5):e0069224. https://doi.org/10.1128/mbio.00692-24.

Callahan BJ, Grinevich D, Thakur S, Balamotis MA, Yehezkel TB. Ultra-accurate microbial amplicon sequencing with synthetic long reads. Microbiome. 2021:**9**(1):130. https://doi.org/10.1186/s40168-021-01072-3.

Caspi I, Meir M, Ben Nun N, Abu Rass R, Yakhini U, Stern A, Ram Y. Mutation rate, selection, and epistasis inferred from RNA virus haplotypes via neural posterior estimation. Virus Evol. 2023:**9**(1):vead033. https://doi.org/10.1093/ve/vead033.

Chamakura KR, Edwards GB, Young R. Mutational analysis of the MS2 lysis protein L. Microbiology. 2017:**163**(7):961–969. https://doi.org/10.1099/mic.0.000485.

Chao L, Elena SF. Nonlinear trade-offs allow the cooperation game to evolve from Prisoner's dilemma to snowdrift. Proc Biol Sci. 2017:**284**(1854):20170228. https://doi.org/10.1098/rspb.2017.0228.

Chaturvedi S, Vasen G, Pablo M, Chen X, Beutler N, Kumar A, Tanner E, Illouz S, Rahgoshay D, Burnett J. Identification of a therapeutic interfering particle—a single-dose SARS-CoV-2 antiviral intervention with a high barrier to resistance. Cell. 2021:**184**(25):6022–6036.e18. https://doi.org/10.1016/j.cell.2021.11.004.

Ciota AT, Ehrbar DJ, Van Slyke GA, Willsey GG, Kramer LD. Cooperative interactions in the west Nile virus mutant swarm. BMC Evol Biol. 2012:**12**(1):58. https://doi.org/10.1186/1471-2148-12-58.

Dai X, Li Z, Lai M, Shu S, Du Y, Zhou ZH, Sun R. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. Nature. 2017:**541**(7635):112–116. https://doi.org/10.1038/nature20589.

Dang M, Wu LJ, Zhang SR, Zhu JR, Hu YZ, Yang CX, Zhang XY. MS2 Virus-like particles as a Versatile peptide presentation platform: insights into the deterministic abilities for accommodating heterologous peptide lengths. ACS Synth Biol. 2023:**12**(12):3704–3715. https://doi.org/10.1021/acssynbio.3c00503.

DaPalma T, Doonan BP, Trager NM, Kasman LM. A systematic approach to virus–virus interactions. Virus Res. 2010:**149**(1):1–9. https://doi.org/10.1016/j.virusres.2010.01.002.

Dennehy JJ, Turner PE. Reduced fecundity is the cost of cheating in RNA virus ϕ 6. Proc Biol Sci. 2004:**271**(1554):2275–2282. https://doi.org/10.1098/rspb.2004.2833.

Díaz-Muñoz SL, Sanjuán R, West S. Sociovirology: conflict, cooperation, and communication among viruses. Cell Host Microbe. 2017:**22**(4):437–441. https://doi.org/10.1016/j.chom.2017.09.012.

Dimmock NJ, Easton AJ. Defective interfering influenza virus RNAs: time to reevaluate their clinical potential as broad-spectrum antivirals? J Virol. 2014:**88**(10):5217–5227. https://doi.org/10.1128/JVI.03193-13.

Duin JV, Tsareva N. Chapter 15: Single-stranded RNA phages. In: Calendar R, Abedon ST, editors. Bacteriophages. New York (NY): Oxford Academic; 2006. p. 175–196. https://doi.org/10.1093/oso/9780195148503.003.0015.

Enea V, Zinder ND. A deletion mutant of bacteriophage f1 containing no intact cistrons. Virology. 1975:**68**(1):105–114. https://doi.org/10.1016/0042-6822(75)90152-X.

Felt SA, Sun Y, Jozwik A, Paras A, Habibi MS, Nickle D, Anderson L, Achouri E, Feemster KA, Cárdenas AM. Detection of respiratory syncytial virus defective genomes in nasal secretions is associated with distinct clinical outcomes. Nat Microbiol. 2021:**6**(5):672–681. https://doi.org/10.1038/s41564-021-00882-3.

Fiegna F, Yu Y-TN, Kadam SV, Velicer GJ. Evolution of an obligate social cheater to a superior cooperator. Nature. 2006:**441**(7091):310–314. https://doi.org/10.1038/nature04677.

Fodor E, Mingay LJ, Crow M, Deng T, Brownlee GG. A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase promotes the generation of defective interfering RNAs. J Virol. 2003:**77**(8):5017–5020. https://doi.org/10.1128/JVI.77.8.5017-5020.2003.

Fueyo R, Judd J, Feschotte C, Wysocka J. Roles of transposable elements in the regulation of mammalian transcription. Nat Rev Mol Cell Biol. 2022:**23**(7):481–497. https://doi.org/10.1038/s41580-022-00457-y.

Gao Y, Abreha M, Nelson KN, Baird H, Dudley DM, Abraha A, Arts EJ. Enrichment of intersubtype HIV-1 recombinants in a dual infection system using HIV-1 strain-specific siRNAs. Retrovirology. 2011:**8**(1):5. https://doi.org/10.1186/1742-4690-8-5.

Gelbart M, Harari S, Ya B-A, Kustin T, Wolf D, Mandelboim M, Mor O, Pennings PS, Stern A. Drivers of within-host genetic diversity in acute infections of viruses. PLoS Pathog. 2020:**16**(11):e1009029. https://doi.org/10.1371/journal.ppat.1009029.

Genoyer E, López CB. The impact of defective viruses on infection and immunity. Annu Rev Virol. 2019:**6**(1):547–566. https://doi.org/10.1146/annurev-virology-092818-015652.

Ghoul M, Griffin AS, West SA. Toward an evolutionary definition of cheating. Evolution. 2014:**68**(2):318–331. https://doi.org/10.1111/evo.12266.

Hall RJ, Whelan FJ, McInerney JO, Ou Y, Domingo-Sananes MR. Horizontal gene transfer as a source of conflict and cooperation in prokaryotes. Front Microbiol. 2020:**11**:1569. https://doi.org/10.3389/fmicb.2020.01569.

Horiuchi K. Co-evolution of a filamentous bacteriophage and its defective interfering particles. J Mol Biol. 1983:**169**(2):389–407. https://doi.org/10.1016/S0022-2836(83)80057-6.

Hunter M, Fusco D. Superinfection exclusion: a viral strategy with short-term benefits and long-term drawbacks. PLoS Comput Biol. 2022:**18**(5):e1010125. https://doi.org/10.1371/journal.pcbi.1010125.

Johansson HE, Liljas L, Uhlenbeck OC. RNA recognition by the MS2 phage coat protein. Semin Virol. 1997:**8**(3):176–185. https://doi.org/10.1006/smvy.1997.0120.

Klinger E, Rickert D, Hasenauer J. pyABC: distributed, likelihood-free inference. Bioinformatics. 2018:**34**(20):3591–3593. https://doi.org/10.1093/bioinformatics/bty361.

Lago H, Fonseca SA, Murray JB, Stonehouse NJ, Stockley PG. Dissecting the key recognition features of the MS2 bacteriophage translational repression complex. Nucleic Acids Res. 1998:**26**(5):1337–1344. https://doi.org/10.1093/nar/26.5.1337.

Leeks A, Bono LM, Ampolini EA, Souza LS, Höfler T, Mattson CL, Dye AE, Díaz-Muñoz SL. Open questions in the social lives of viruses. J Evol Biol. 2023a:**36**(11):1551–1567. https://doi.org/10.1111/jeb.14203.

Leeks A, Sanjuán R, West SA. The evolution of collective infectious units in viruses. Virus Res. 2019:**265**:94–101. https://doi.org/10.1016/j.virusres.2019.03.013.

Leeks A, West SA, Ghoul M. The evolution of cheating in viruses. Nat Commun. 2021:**12**(1):6928. https://doi.org/10.1038/s41467-021-27293-6.

Leeks A, Young PG, Turner PE, Wild G, West SA. Cheating leads to the evolution of multipartite viruses. PLoS Biol. 2023b:**21**(4):e3002092. https://doi.org/10.1371/journal.pbio.3002092.

Lintusaari J, Gutmann MU, Dutta R, Kaski S, Corander J. Fundamentals and recent developments in approximate Bayesian computation. Syst Biol. 2017:**66**(1):e66–e82. https://doi.org/10.1093/sysbio/syw077.

McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004:**32**(Web Server):W20–W25. https://doi.org/10.1093/nar/gkh435.

Meir M, Harel N, Miller D, Gelbart M, Eldar A, Gophna U, Stern A. Competition between social cheater viruses is driven by mechanistically different cheating strategies. Sci Adv. 2020:**6**(34):eabb7990. https://doi.org/10.1126/sciadv.abb7990.

Mezhyrova J, Martin J, Börnsen C, Dötsch V, Frangakis AS, Morgner N, Bernhard F. In vitro characterization of the phage lysis protein MS2-L. Microbiome Res Rep. 2023:**2**(4):28. https://doi.org/10.20517/mrr.2023.28.

Naskalska A, Heddle JG. Virus-like particles derived from bacteriophage MS2 as antigen scaffolds and RNA protective shells. Nanomedicine (Lond). 2024:**19**(12):1103–1115. https://doi.org/10.2217/nnm-2023-0362.

Pelz L, Rüdiger D, Dogra T, Alnaji FG, Genzel Y, Brooke CB, Kupke SY, Reichl U. Semi-continuous propagation of influenza A virus and its defective interfering particles: analyzing the dynamic competition to select candidates for antiviral therapy. J Virol. 2021:**95**(24):e0117421. 10.1128/jvi. 01174-01121.

Ranum JN, Ledwith MP, Alnaji FG, Diefenbacher M, Orton R, Sloan E, Güereca M, Feltman EM, Smollett K, da Silva Filipe A. Cryptic proteins translated from deletion-containing viral genomes dramatically expand the influenza virus proteome. Nucleic Acids Res. 2024:**52**(6):3199–3212. https://doi.org/10.1093/nar/gkae133.

Rast LI, Rouzine IM, Rozhnova G, Bishop L, Weinberger AD, Weinberger LS. Conflicting selection pressures will constrain viral escape from interfering particles: principles for designing resistance-proof antivirals. PLoS Comput Biol. 2016:**12**(5):e1004799. https://doi.org/10.1371/journal.pcbi.1004799.

Rezelj VV, Carrau L, Merwaiss F, Levi LI, Erazo D, Tran QD, Henrion-Lacritick A, Gausson V, Suzuki Y, Shengjuler D. Defective viral genomes as therapeutic interfering particles against flavivirus infection in mammalian and mosquito hosts. Nat Commun. 2021:**12**(1):2290. https://doi.org/10.1038/s41467-021-22341-7.

Rith S, Chin S, Sar B, Phalla Y, Horm SV, Ly S, Buchy P, Dussart P, Horwood PF. Natural co-infection of influenza A/H3N2 and A/H1N1pdm09 viruses resulting in a reassortant A/H3N2 virus. J Clin Virol. 2015:**73**:108–111. https://doi.org/10.1016/j.jcv.2015.11.008.

Rolfsson Ó, Middleton S, Manfield IW, White SJ, Fan B, Vaughan R, Ranson NA, Dykeman E, Twarock R, Ford J. Direct evidence for packaging signal-mediated assembly of bacteriophage MS2. J Mol Biol. 2016:**428**(2):431–448. https://doi.org/10.1016/j.jmb.2015.11.014.

Sanjuán R. Collective infectious units in viruses. Trends Microbiol. 2017:**25**(5):402–412. https://doi.org/10.1016/j.tim.2017.02.003.

Sanjuán R, Thoulouze M-I. Why viruses sometimes disperse in groups. Virus Evol. 2019:**5**(1):vez014. https://doi.org/10.1093/ve/vez014.

Santiana M, Ghosh S, Ho BA, Rajasekaran V, Du W-L, Mutsafi Y, De Jésus-Diaz DA, Sosnovtsev SV, Levenson EA, Parra GI. Vesicle-cloaked virus clusters are optimal units for inter-organismal viral transmission. Cell Host Microbe. 2018:**24**(2):208–220.e8. https://doi.org/10.1016/j.chom.2018.07.006.

Schmidt BF, Berkhout B, Overbeek GP, van Strien A, van Duin J. Determination of the RNA secondary structure that regulates lysis gene expression in bacteriophage MS2. J Mol Biol. 1987:**195**(3):505–516. https://doi.org/10.1016/0022-2836(87)90179-3.

Secor PR, Dandekar AA. More than simple parasites: the sociobiology of bacteriophages and their bacterial hosts. mBio. 2020:**11**(2):e00041–e00020. 10.1128/mbio. 00041-00020.

Segredo-Otero E, Sanjuán R. Cooperative virus-virus interactions: an evolutionary perspective. Biodes Res. 2022:**2022**:9819272. https://doi.org/10.34133/2022/9819272.

Simon B, Fletcher JA, Doebeli M. Towards a general theory of group selection. Evolution. 2013:**67**(6):1561–1572. https://doi.org/10.1111/j.1558-5646.2012.01835.x.

Singhal S, Turner PE. Effects of historical co-infection on host shift abilities of exploitative and competitive viruses. Evolution. 2021:**75**(7):1878–1888. https://doi.org/10.1111/evo.14263.

Stonehouse NJ, Valegård K, Golmohammadi R, van den Worm S, Walton C, Stockley PG, Liljas L. Crystal structures of MS2 capsids with mutations in the subunit FG loop. J Mol Biol. 1996:**256**(2):330–339. https://doi.org/10.1006/jmbi.1996.0089.

Tanner EJ, Jung S-Y, Glazier J, Thompson C, Zhou Y, Martin B, Son H-I, Riley JL, Weinberger LS. Discovery and Engineering of a Therapeutic Interfering Particle (TIP): a combination self-renewing antiviral. bioRxiv 820456. https://doi.org/10.1101/029983. 30 October 2019, preprint: not peer reviewed.

Traulsen A, Nowak MA. Evolution of cooperation by multilevel selection. Proc Natl Acad Sci U S A. 2006:**103**(29):10952–10955. https://doi.org/10.1073/pnas.0602530103.

Van Himbergen J, Van Geffen B, Van Duin J. Translational control by a long range RNA–RNA interaction; a basepair substitution analysis. Nucleic Acids Res. 1993:**21**(8):1713–1717. https://doi.org/10.1093/nar/21.8.1713.

Vignuzzi M, López CB. Defective viral genomes are key drivers of the virus–host interaction. Nat Microbiol. 2019:**4**(7):1075–1087. https://doi.org/10.1038/s41564-019-0465-y.

Wu H, Zhou H-Y, Zheng H, Wu A. Towards understanding and identification of human viral co-infections. Viruses. 2024:**16**(5):673. https://doi.org/10.3390/v16050673.

Wu M, Zhou E, Sheng R, Fu X, Li J, Jiang C, Su W. Defective interfering particles of influenza virus and their characteristics, impacts, and use in vaccines and antiviral strategies: a systematic review. Viruses. 2022:**14**(12):2773. https://doi.org/10.3390/v14122773.

Xu J, Sun Y, Li Y, Ruthel G, Weiss SR, Raj A, Beiting D, López CB. Replication defective viral genomes exploit a cellular pro-survival mechanism to establish paramyxovirus persistence. Nat Commun. 2017:**8**(1):799. https://doi.org/10.1038/s41467-017-00909-6.